# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-318905
(43)Date of publication of application : 16.11.2001

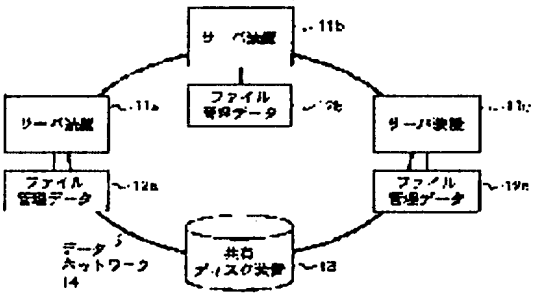
(51)Int.Cl. G06F 15/177
G06F 3/06
G06F 12/00
G06F 15/00
G06F 15/167

(21)Application number : 2000-133585
(22)Date of filing : 02.05.2000

(71)Applicant : MATSUSHITA ELECTRIC IND CO LTD
(72)Inventor : NAKATSUKA MONTA

## (54) DISK SHARED TYPE DISTRIBUTED SERVER SYSTEM

(57)Abstract:
PROBLEM TO BE SOLVED: To solve problems that the installation cost of another mechanism for executing exclusive control is required though each server device executes the exclusive control by the mechanism existing in a place (e.g. a device annexed to a shared disk or a completely different device) different from the server device concerned in a distributed file system, and when a fault is generated in the other mechanism, the whole system is stopped.
SOLUTION: The system for sharing a disk by plural server devices divides information written in a file management data and each server device independently manages the information. Only when a request is not satisfied only by the file management data stored in the server device itself, the server device inquires the other server device whether the device has the information of the file management data or not.




LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2001－318905
（P2001－318905A）
(43)公開日　平成13年11月16日(2001.11.16)

| (51)Int.Cl.' | | 識別記号 | FI | | テーマコード'(参考) |
|---|---|---|---|---|---|
| Ｇ０６Ｆ | 15/177 | ６８２ | Ｇ０６Ｆ 15/177 | ６８２Ｂ | ５Ｂ０４５ |
| | 3/06 | ３０１ | 3/06 | ３０１Ｊ | ５Ｂ０６５ |
| | 12/00 | ５２０ | 12/00 | ５２０Ｊ | ５Ｂ０８２ |
| | | ５４５ | | ５４５Ａ | ５Ｂ０８５ |
| | 15/00 | ３１０ | 15/00 | ３１０Ａ | |

審査請求　未請求　請求項の数９　ＯＬ　（全９頁）　最終頁に続く

(21)出願番号　特願2000－133585(P2000－133585)

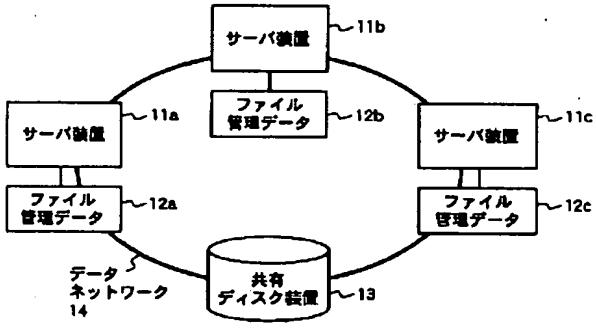
(22)出願日　平成12年５月２日(2000.5.2)

(71)出願人　000005821
松下電器産業株式会社
大阪府門真市大字門真1006番地
(72)発明者　中塚　紋太
大阪府門真市大字門真1006番地　松下電器産業株式会社内
(74)代理人　100081813
弁理士　早瀬　憲一

Ｆターム(参考)　5B045 BB12 BB28 BB47 DD03 DD16
EE03 EE25 GG01
5B065 BA01 CA02 CC03 CC08 CE26
ZA01 ZA08 ZA15
5B082 CA08 HA08 HA09
5B085 BA06 BE07 BG04 BG07

(54)【発明の名称】　ディスク共有型分散サーバシステム

(57)【要約】
【課題】　分散型ファイルシステムでは、サーバ装置とは別の場所（例えば、共有ディスクに付属した装置、または全く別の装置）において、排他制御を司る機構が存在し、各々のサーバ装置はその機構に頼って排他制御を行う。これによりファイルアクセスの負荷は分散されるが、排他制御を司る別機構を設けるコストがかかる。また上記別機構に障害が発生した場合、システム全体が停止してしまうなどの問題があった。
【解決手段】　複数のサーバ装置でディスクを共有するようなシステムにおいて、ファイル管理データに書かれた情報を分割し、サーバ装置はその情報を独立して管理する。サーバ装置は、クライアントからの要求に従い、自らが持つファイル管理データだけでは要求が満たせない場合のみ、他のサーバ装置にファイル管理データの情報を問い合わせる。



1

【特許請求の範囲】
【請求項１】　データの入出力を目的とした複数サーバ装置が、データネットワークを介して、データを格納するためのディスク装置を共有する構成をとるディスク共有型分散サーバシステムにおいて、
前記ディスク装置におけるデータを論理的なファイルとして扱うためのファイル管理データを、予め２つ以上の情報に分割しておき、前記分割された情報を、各サーバ装置内部において管理するファイルシステムを備え、
前記複数サーバ装置のうちの１つがこれに接続されるクライアントからのデータ配信要求を受けると、該サーバ装置のファイルシステムは、管理するファイル管理データの情報の中から該当するコンテンツファイルを検索し、該当コンテンツファイルが存在した場合、前記ディスク装置に対してアクセスを行ない、読み出しを行う、
ことを特徴とするディスク共有型分散サーバシステム。
【請求項２】　請求項１記載のディスク共有型分散サーバシステムにおいて、
前記該当コンテンツファイルが存在しなかった場合、前記ファイルシステムは、前記データネットワークを介して接続されている他のサーバ装置に対して、前記該当コンテンツファイルが存在するかどうかを確認し、前記該当コンテンツファイルの情報を持つ前記他のサーバ装置よりデータの読み出しに必要な情報を受け取る手段を備え、
前記ディスク装置に対してアクセスを行い、読み出しを行う、
ことを特徴とするディスク共有型分散サーバシステム。
【請求項３】　請求項１または請求項２記載のディスク共有型分散サーバシステムにおいて、
前記複数サーバ装置のうちの１つがこれに接続されるクライアントからデータ記録要求を受けると、ファイルシステムは前記データネットワークを介して接続されている、他のサーバ装置が管理するファイル管理データの情報の中に、記録要求されているコンテンツファイルと同一名のものがない事を確認する手段と、
前記ファイルシステムが管理するファイル管理データの情報の中から、記録するのに必要なディスク容量を確保する手段を備え、
前記ディスク装置に対して書き込みを行う、
ことを特徴とするディスク共有型分散サーバシステム。
【請求項４】　請求項３記載のディスク共有型分散サーバシステムにおいて、
前記ファイルシステムがファイル管理データを参照して記録するのに必要なディスク容量が確保できなかった場合、
前記ファイルシステムは、前記データネットワークを介して接続されている、他のサーバ装置内部にあるファイルシステムに対してコンテンツファイルの情報の管理を要求する手段と、

2

前記ディスク装置に対して書き込む際に、前記コンテンツファイルの情報の管理を要求した他のサーバ装置よりファイルアクセス（書き込み）に必要な情報を受け取る手段を備え、
前記ディスク装置に対して書き込みを行う、
ことを特徴とするディスク共有型分散サーバシステム。
【請求項５】　請求項３または請求項４記載のディスク共有型分散サーバシステムにおいて、
各サーバ装置内部のファイルシステムは、共有している前記ディスク装置に記録されている全コンテンツファイル名および前記コンテンツファイルの情報を管理しているサーバ装置を対応付けさせた管理テーブルを持つ、
ことを特徴とするディスク共有型分散サーバシステム。
【請求項６】　請求項５記載のディスク共有型分散サーバシステムにおいて、
コンテンツファイルの追加または削除を行ったサーバ装置は、前記データネットワークを介して接続されている他のサーバ装置に対して、前記管理テーブルの更新を指示する手段を備えた、
ことを特徴とするディスク共有型分散サーバシステム。
【請求項７】　請求項５または請求項６記載のディスク共有型分散サーバシステムにおいて、
各サーバ装置内部のファイルシステムは、前記分割した記録領域における空き情報を空き領域テーブルとして持つ、
ことを特徴とするディスク共有型分散サーバシステム。
【請求項８】　請求項７記載のディスク共有型分散サーバシステムにおいて、
前記空き領域情報の更新を行ったサーバ装置は、前記データネットワークを介して接続されている他のサーバ装置に対して、前記空き領域テーブルの更新を指示する手段を備えた、
ことを特徴とするディスク共有型分散サーバシステム。
【請求項９】　請求項３または請求項４記載のディスク共有型分散サーバシステムにおいて、
システム初期化時に、前記ディスク装置における記録領域を論理的に分割する手段と、
前記分割した記録領域に内在するコンテンツファイルの情報を取得する手段とを備えた、
ことを特徴とするディスク共有型分散サーバシステム。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、データの入出力を目的とした複数サーバ装置が、データネットワークを介して、データを格納するためのディスク装置を共有する構成をとるディスク共有型分散サーバシステムに関するものである。
【０００２】
【従来の技術】サーバ装置とは、例えばクライアントからのデータ配信要求に応じて、ディスク装置からデータ

を読み出しクライアントに送信するというものである。このサーバ装置内部において、ディスク上に格納したデータの集合を論理的なファイルとして扱うため、独自のファイル管理データによりファイルを管理するのが、ファイルシステムと呼ばれるものである。

【０００３】このようなサーバ装置を分散型モデルによって実現する上では、複数のサーバ装置間で、このファイル管理データの排他制御を如何にして行うかが課題であった。

【０００４】従来では、あるサーバ装置内部に排他制御を司る機構がファイル管理データを一括して管理することで、システム全体のファイルオペレーションの承認を１つのサーバ装置が行っていた。この方式では実装やデータの一貫性という面に優れている反面、サーバ装置の台数が増える事により排他制御機構によるボトルネックが発生してしまい、また排他制御を行うサーバ装置に故障が発生するとシステム全体に影響を与えてしまう。

【０００５】例えば特開平１１－１４６３４３号公報に開示されたディスク共有システムに記載されているように、排他制御を司る機構をシステムの一部に設け、ファイルオペレーションに関する承認を一手に引き受ける事で、ファイル管理データの一貫性を保つというものである。

【０００６】

【発明が解決しようとする課題】しかしながらこのような方式では、排他制御をサーバ装置とは別機構にすることでコスト面の負荷があり、また排他制御機構がシステムの弱点となり、そこで故障が発生するとシステム全体に影響を与えることとなる。このため、排他制御機構部には、信頼性向上のための機構が別途必要となり、さらにコストの負荷が増える。

【０００７】本発明は、ディスク装置を共有する構成をとるディスク共有型分散サーバシステムにおいて、特定のサーバ装置に対して負荷を集中させることなく、またコストを低く抑える事ができ、かつ障害発生時にシステム全体に致命的な影響を与えることのないファイルシステムである、ディスク共有型分散サーバシステムの提供を目的としている。

【０００８】

【課題を解決するための手段】この課題を解決するために、請求項１に記載のディスク共有型分散サーバシステムは、データの入出力を目的とした複数サーバ装置が、データネットワークを介して、データを格納するためのディスク装置を共有する構成をとるディスク共有型分散サーバシステムにおいて、前記ディスク装置におけるデータを論理的なファイルとして扱うためのファイル管理データを、予め２つ以上の情報に分割しておき、前記分割された情報を、各サーバ装置内部において管理するファイルシステムを備え、前記複数サーバ装置のうちの１つがこれに接続するクライアントからのデータ配信要求を受けると、該サーバ装置のファイルシステムは、管理するファイル管理データの情報の中から該当するコンテンツファイルを検索し、該当コンテンツファイルが存在した場合、前記ディスク装置に対してアクセスを行ない、読み出しを行う、ことを特徴とする。



【０００９】請求項２に記載のディスク共有型分散サーバシステムは、請求項１記載のディスク共有型分散サーバシステムにおいて、前記該当コンテンツファイルが存在しなかった場合、前記ファイルシステムは、前記データネットワークを介して接続されている他のサーバ装置に対して、前記該当コンテンツファイルが存在するかどうかを確認し、前記該当コンテンツファイルの情報を持つ前記他のサーバ装置よりデータの読み出しに必要な情報を受け取る手段を備え、前記ディスク装置に対してアクセスを行い、読み出しを行う、ことを特徴とする。

【００１０】請求項３に記載のディスク共有型分散サーバシステムは、請求項１または請求項２記載のディスク共有型分散サーバシステムにおいて、前記複数サーバ装置のうちの１つがこれに接続するクライアントからデータ記録要求を受けると、該サーバ装置のファイルシステムは、前記データネットワークを介して接続されている、他のサーバ装置が管理するファイル管理データの情報の中に、記録要求されているコンテンツファイルと同一名のものがない事を確認する手段と、自分が管理するファイル管理データの情報の中から、記録するのに必要なディスク容量を確保する手段を備え、前記ディスク装置に対して書き込みを行う、ことを特徴とする。

【００１１】請求項４に記載のディスク共有型分散サーバシステムは、請求項３記載のディスク共有型分散サーバシステムにおいて、前記ファイルシステムがファイル管理データを参照して記録するのに必要なディスク容量が確保できなかった場合、前記ファイルシステムは、前記データネットワークを介して接続されている、他のサーバ装置内部にあるファイルシステムに対してコンテンツファイルの情報の管理を要求する手段と、前記ディスク装置に対して書き込む際に、前記コンテンツファイルの情報の管理を要求した他のサーバ装置よりファイルアクセス（書き込み）に必要な情報を受け取る手段を備え、前記ディスク装置に対して書き込みを行う、ことを特徴とする。

【００１２】請求項５に記載のディスク共有型分散サーバシステムは、請求項３または請求項４記載のディスク共有型分散サーバシステムにおいて、各サーバ装置内部のファイルシステムは、共有している前記ディスク装置に記録されている全コンテンツファイル名および前記コンテンツファイルの情報を管理しているサーバ装置を対応付けさせた管理テーブルを持つ、ことを特徴とする。

【００１３】請求項６に記載のディスク共有型分散サーバシステムは、請求項５記載のディスク共有型分散サーバシステムにおいて、コンテンツファイルの追加または

削除を行ったサーバ装置は、前記データネットワークを介して接続されている他のサーバ装置に対して、前記管理テーブルの更新を指示する手段を備えた、ことを特徴とする。

【００１４】請求項７に記載のディスク共有型分散サーバシステムは、請求項５または請求項６記載のディスク共有型分散サーバシステムにおいて、各サーバ装置内部のファイルシステムは、前記分割した記録領域における空き情報を空き領域テーブルとして持つ、ことを特徴とする。

【００１５】請求項８に記載のディスク共有型分散サーバシステムは、請求項７記載のディスク共有型分散サーバシステムにおいて、前記空き領域情報の更新を行ったサーバ装置は、前記データネットワークを介して接続されている他のサーバ装置に対して、前記空き領域テーブルの更新を指示する手段を備えた、ことを特徴とする。

【００１６】請求項９に記載のディスク共有型分散サーバシステムは、請求項３または請求項４記載のディスク共有型分散サーバシステムにおいて、システム初期化時に、前記ディスク装置における記録領域を論理的に分割する手段と、前記分割した記録領域に内在するコンテンツファイルの情報を取得する手段を備えた、ことを特徴とする。

【００１７】上記のような構成によって、各ファイルの管理データ、コンテンツテーブル、空き領域の情報を複数のサーバ装置に分散させたことにより、１つのサーバ装置に障害が発生した場合でも、他のサーバ装置が管理するディスクエリアのコンテンツファイルの入出力に関しては支障なく動作することが可能となる。また、システム構成を変更した場合でも、システムを再立ち上げすることにより、他サーバ装置のディスク情報を容易に持つことができる、という作用を有する。

【００１８】

【発明の実施の形態】以下、本発明にかかるディスク共有型分散サーバシステムの実施例について、図を用いて説明する。

（実施の形態１）以下に本発明の請求項１ないし請求項８に対応する実施の形態１について説明する。図１は、本発明の実施の形態１におけるディスク共有型分散サーバシステムの構成図を示したものである。図１では３つのサーバ装置が接続されているシステムを例としている。

【００１９】図中、１１ａ，１１ｂ，１１ｃは、クライアントからビデオデータなどの記録再生要求を受け取って動作するサーバ装置Ａ，Ｂ，Ｃを示し、ハードウェア構成およびソフトウェア構成を全く同じとする。１３はビデオデータを含むコンテンツが格納されている共有ディスク装置である。各サーバ装置１１ａ，１１ｂ，１１ｃと共有ディスク装置１３とは、Ｆｉｂｒｅ　Ｃｈａｎｎｅｌに代表されるような複数のプロトコルによるデータの送受信が可能なデータネットワーク１４で接続されている。また、各サーバ装置１１ａ，１１ｂ，１１ｃ内部のファイルシステムは、共有ディスク装置１３上に格納したデータの集合を論理的なファイルとして扱うため、それぞれファイル管理データＡ１２ａ，ファイル管理データＢ１２ｂ，ファイル管理データＣ１２ｃを持っており、共有ディスク装置１３に対してアクセスすることが可能なようになっている。

【００２０】次にファイル管理データの構成を図２において示す。図中、１２ａ，１２ｂ，１２ｃは、各サーバ装置内部のファイルシステムが管理しているファイル管理データＡ，Ｂ，Ｃのイメージ図である。各ファイル管理データ１２ａ，１２ｂ，１２ｃは、主に２つの情報から構成されている。１つ目は共有ディスク装置１３の記録領域の状態（使用／未使用）を示すディスク領域情報であり、共有ディスク装置１３を論理的に３つのエリアＡ，Ｂ，Ｃに分割し、それぞれ管理エリアＡのディスク領域情報２１ａ、管理エリアＢのディスク領域情報２１ｂ、管理エリアＣのディスク領域情報２１ｃで示している。これらのディスク領域情報は互いにエリア重複していない。２つ目はディスク領域情報２２ａ，２２ｂ，２２ｃに示されている記録領域内に存在しているコンテンツの情報であり、２２ａ，２２ｂ，２２ｃはぞれぞれ、エリアＡに含まれているコンテンツファイル１～ｐ情報、エリアＢに含まれているコンテンツファイルｐ＋１～ｑ情報、エリアＣに含まれているコンテンツファイルｑ＋１～ｒ情報を示している。

【００２１】次に図１の構成において、システムにおける動作を、図３及び図４を参照して説明する。各図において、太字破線矢印は制御またはデータアクセス、太字実線矢印はデータの流れを示す。図３は、サーバ装置がこれに接続されているクライアントからコンテンツファイルｍ（１≦ｍ≦ｐ）の配信要求を受けた場合の動作を示す説明図である。サーバ装置Ａ１１ａが接続されているクライアントからコンテンツファイルｍ（１≦ｍ≦ｐ）の配信要求３１を受け取ったとする。サーバ装置Ａ１１ａのファイルシステムは、ファイル管理データＡ１２ａを参照してコンテンツファイルｍの情報検索３２を行う。該当したコンテンツファイルｍ情報を元に共有ディスク装置１３からデータ読み出し３３を行ない、所定の出力先へデータ配信３４を行う。

【００２２】図４は、サーバ装置がこれに接続されているクライアントからコンテンツファイルｎ（ｐ＋１≦ｎ≦ｑ）の配信要求を受けた場合の動作を示す説明図である。サーバ装置Ａ１１ａが接続されているクライアントからコンテンツファイルｎ（ｐ＋１≦ｎ≦ｑ）の配信要求４１を受け取ったとする。サーバ装置Ａ１１ａのファイルシステムはファイル管理データＡ１２ａを参照してコンテンツファイルｎの情報検索４２を行うが、該当情報が見つからないので、サーバ装置Ｂ１１ｂに対してデ

ータネットワーク１４を通じて情報検索の依頼４３を行う。サーバ装置Ｂ１１ｂのファイルシステムはファイル管理データＢ１２ｂを参照してコンテンツファイルｎの情報検索４４を行う。該当したコンテンツファイルｎ情報をサーバ装置Ａ１１ａに送信４５することによりサーバ装置Ａ１１ａは共有ディスク装置１３からデータの読み出し４６を行い、所定の出力先へデータ配信４７を行う。上記の方法の場合、コンテンツファイルｎの検索を、接続されているサーバ装置全部に対して問い合わせしなければならない状況が発生するため、さらに効率よく検索できるよう、図５に示すようなテーブルを用意する。

【００２３】図５は、コンテンツテーブルの内容を示す説明図である。図５に示すように、コンテンツテーブルはコンテンツファイル名に相対して、状態および保存場所を示している。これは図１で示した構成のサーバ装置Ａ，Ｂ，Ｃにおいて、各々が同期を取って同じ内容のテーブルを持つこととする。このテーブルの更新（コンテンツファイルの追加および削除）を行うサーバ装置は、他のサーバ装置に対して情報の更新を指示することで、データ同期を保つ。

【００２４】図６は、図１の構成に図５で示したコンテンツテーブルを備えたディスク共有型分散サーバシステムの構成を示す説明図である。図６において、１１ａ、１１ｂ、１１ｃはサーバ装置、１２ａ、１２ｂ、１２ｃはファイル管理データ、１３は共有ディスク装置、１４はデータネットワークである。また、６０ａ、６０ｂ、６０ｃは、それぞれサーバ装置１１ａ、１１ｂ、１１ｃに含まれるコンテンツテーブルである。次に図６の構成において、システムにおける動作を、図７、図８及び図９を参照して説明する。各図において、太字破線矢印は制御またはデータアクセス、太字実線矢印はデータの流れを示す。

【００２５】図６は、サーバ装置がこれに接続されているクライアントからコンテンツファイルｎ（ｐ＋１≦ｎ≦ｑ）の配信要求を受けた場合の動作を示す説明図である。サーバ装置Ａ１１ａが接続されているクライアントからコンテンツファイルｎ（ｐ＋１≦ｎ≦ｑ）の配信要求６１を受け取ったとする。サーバ装置Ａ１１ａのファイルシステムはコンテンツテーブル６０ａを元にして該当コンテンツファイルがサーバ装置Ｂ１１ｂの管理下であることを確認し、サーバ装置Ｂ１１ｂに対してデータネットワーク１４を通じて情報検索の依頼６２を行う。サーバ装置Ｂ１１ｂのファイルシステムは、ファイル管理データＢ１２ｂを参照してコンテンツファイルｎの情報取得６３を行い、ファイルｎ情報をサーバ装置Ａ１１ａに送信６４することにより、サーバ装置Ａ１１ａは共有ディスク装置１３からデータの読み出し６５を行い、所定の出力先へデータ配信６６を行う。

【００２６】図７は、図５で示したコンテンツテーブルを備えたサーバ装置が、これに接続されているクライアントから、新しいコンテンツファイルｘの記録要求を受けた場合の動作を示す説明図である。図７において、サーバ装置Ａ１１ａが接続されているクライアントから新しいコンテンツファイルｘの記録要求７１を受け取ったとする。サーバ装置Ａ１１ａのファイルシステムはコンテンツテーブル６０ａにおいて同一のコンテンツファイル名が存在しない事を確認した後、ファイル管理データＡ１２ａにおいて、共有ディスク装置１３の記録領域エリアＡに空き領域が存在する事を確認して、新しいコンテンツファイルｘをコンテンツテーブル６０ａに追加する。次に所定の入力先よりデータ受信７３を行ない、記録場所をファイル管理データＡ１２ａにおけるコンテンツファイルｘの情報に書き込み７４、共有ディスク装置１３にデータ書き込み７５を行う。

【００２７】図８は、図５で示したコンテンツテーブルを備えたサーバ装置が、これに接続されているクライアントから新しいコンテンツファイルの記録要求を受け、他のサーバ装置のファイル管理データを用いる場合の動作を示す説明図である。図８において、サーバ装置Ａ１１ａが接続されているクライアントから新しいコンテンツファイルｘの記録要求８１を受け取ったとする。サーバ装置Ａ１１ａのファイルシステムはコンテンツテーブルにおいて同一のコンテンツファイル名が存在しない事を確認した後、ファイル管理データＡ１２ａにおいて、共有ディスク装置１３の記録領域エリアＡに空き領域が存在しないと判断８２した場合、サーバ装置Ｂ１１ｂに対してデータネットワーク１４を通じてエリアＢの空き領域の確認依頼８３を行う。サーバ装置Ｂ１１ｂのファイルシステムは、ファイル管理データＢ１２ｂを参照して、エリアＢの空き領域の確認８４を行うと共に、新しいコンテンツファイルｘをコンテンツテーブル６０ｂに追加する。次に空き領域情報つまり共有ディスクにおいて書き込み可能な物理領域の情報をサーバ装置Ａ１１ａに送信８５することにより、サーバ装置Ａ１１ａは所定の入力先よりデータ受信８６を行い、記録場所をファイル管理データＢ１２ｂにおけるコンテンツファイルｘの情報に書き込み８７、共有ディスク装置１３にデータ書き込み８８を行う。

【００２８】図９は、図１の構成に図５で示したコンテンツテーブルと、後述する領域テーブルとを備えたディスク共有型分散サーバシステムの構成を示す説明図である。図９において、１１ａ、１１ｂ、１１ｃはサーバ装置、１２ａ、１２ｂ、１２ｃはファイル管理データ、１３は共有ディスク装置、１４はデータネットワーク、６０ａ、６０ｂ、６０ｃはコンテンツテーブルである。また、９０ａ、９０ｂ、９０ｃは、共有ディスク装置１３における空き容量をエリア別に区分している領域テーブルである。これは図１で示した構成のサーバ装置Ａ，Ｂ，Ｃにおいて、各々が同期を取って同じ内容のテーブ

ルを持つこととする。このテーブルの更新（空き領域の増減）を行うサーバ装置は、他のサーバ装置に対して情報の更新を指示することで、データ同期を保つ。

【００２９】次に動作について図９を参照しながら説明する。図において、太字破線矢印は制御またはデータアクセス、太字実線矢印はデータの流れを示す。サーバ装置Ａ１１ａに接続されているクライアントから新しいコンテンツファイルｘの記録要求９１を受け取ったとする。サーバ装置Ａ１１ａのファイルシステムは、図５で示したコンテンツテーブルにおいて、同一のコンテンツファイル名が存在しない事を確認した上で、領域テーブル９０ａを参照し、予め設定された優先度に従って空き領域のあるエリアを確定し、そのエリアを管理するサーバ装置を特定する。仮にそれがサーバ装置Ｂ１１ｂであった場合、データネットワーク１４を通じて空き領域情報の要求９２を送信し、サーバ装置Ｂ１１ｂは空き領域情報を取得９３した上でサーバ装置Ａ１１ａに情報を送信９４する。サーバ装置Ａ１１ａは所定の入力先よりデータ受信９５を行い、記録場所をファイル管理データＢ１２ｂにおけるコンテンツファイルｘの情報に書き込む９６と同時に、共有ディスク装置１３にデータ書き込み９７を行う。

【００３０】以上のように、本発明の実施の形態１に係るディスク共有型分散サーバシステムによれば、システム固有の排他制御を司る機関の存在しない本ディスク共有型分散サーバシステムにおいて、各ファイルの管理データを複数のサーバ装置に分散させ、複数のサーバ装置が同期を保っているコンテンツテーブルと、領域テーブルとを各サーバ装置に備えたことにより、１つのサーバ装置に障害が発生した場合、該サーバ装置が管理するディスクエリアに記録されているコンテンツファイルの入出力は停止してしまうが、他のサーバ装置が管理するディスクエリアのコンテンツファイルの入出力に関しては全く支障無く動作することが可能となり、コンテンツファイルまたは空き領域の情報の検索を、接続されているサーバ装置全部に対して問い合わせすることなく行うことができる。

【００３１】（実施の形態２）以下に本発明の請求項９に対応する実施の形態２について説明する。図１０は、本発明の実施の形態２に係るディスク共有型分散サーバシステムの構成を示したものである。図中、１０１－１～１０１－ｎは、ディスク一体型のサーバ装置である。これらサーバ装置およびディスク装置１０１－１～１０１－ｎは、Ｆｉｂｒｅ　Ｃｈａｎｎｅｌに代表されるような複数のプロトコルによるデータの送受信が可能なデータネットワーク１０２で接続されている。また、各サーバ装置１０１－１～１０１－ｎ内部のファイルシステムは、一体となっているディスク装置上に格納したデータの集合を論理的なファイルとして扱うためのファイル管理データを、それぞれが持っている。



【００３２】この構成における動作について以下に説明する。システム起動時、つまりサーバ装置およびディスク装置１０１－１～１０１－ｎの全ての電源が投入された状態において、各サーバ装置は、データネットワーク１０２に接続されている他のサーバ装置に対して、各々が管理しているコンテンツの一覧、及び各々が管理しているディスク容量、の２つについて問い合わせを行う。問い合わせを受けたサーバ装置は、ファイル管理テーブルよりその情報を取得して返答する。全ての返答を受け取ると、図５で示したようなコンテンツテーブル、および図９で示した領域テーブルを構築する事で、他のサーバ装置のディスク情報を持つ事になる。

【００３３】以上のように、本発明の実施の形態２に係るディスク共有型分散サーバシステムによれば、上記のような初期化手順を持つ事により、障害が発生した後にシステム構成を変更した場合でも、システムを再立ち上げすることにより、他サーバ装置のディスク情報を容易に持つことが可能である。また、ディスク容量の拡張や、入出力ポートの拡張などの目的により、サーバ装置およびディスク装置の台数を増やす場合でも、システムを再立ち上げすることにより、他サーバ装置のディスク情報を容易に持つことが可能である。

【００３４】

【発明の効果】以上のように、本発明の請求項１および請求項２に係るディスク共有型分散サーバシステムによれば、データの入出力を目的とした複数サーバ装置が、データネットワークを介して、データを格納するためのディスク装置を共有する構成をとるディスク共有型分散サーバシステムにおいて、前記ディスク装置におけるデータを論理的なファイルとして扱うためのファイル管理データを、予め２つ以上の情報に分割しておき、前記分割された情報を、各サーバ装置内部において管理するファイルシステムを備え、前記複数サーバ装置のうちの１つがこれに接続するクライアントからのデータ配信要求を受けると、該サーバ装置のファイルシステムは、管理するファイル管理データの情報の中から該当するコンテンツファイルを検索し、該当コンテンツファイルが存在した場合、前記ディスク装置に対してアクセスを行ない、また、前記該当コンテンツファイルが存在しなかった場合、前記ファイルシステムは、前記データネットワークを介して接続されている他のサーバ装置に対して、前記該当コンテンツファイルが存在するかどうかを確認し、前記該当コンテンツファイルの情報を持つ前記他のサーバ装置よりデータの読み出しに必要な情報を受け取る手段を備え、前記ディスク装置に対してアクセスを行い、読み出しを行うので、１つのサーバ装置に障害が発生した場合でも、該サーバ装置が管理するディスクエリアに記録されているコンテンツファイルの読み出し処理は停止するものの、他のサーバ装置が管理するディスクエリアのコンテンツファイルの読み出しに関しては全く

支障無く動作することが可能となる、という効果を得られる。

【００３５】また、請求項３および請求項４に係るディスク共有型分散サーバシステムによれば、請求項１または請求項２記載のディスク共有型分散サーバシステムにおいて、前記複数サーバ装置のうちの１つがこれに接続するクライアントからデータ記録要求を受けると、該サーバ装置のファイルシステムは、前記データネットワークを介して接続されている、他のサーバ装置が管理するファイル管理データの情報の中に、記録要求されているコンテンツファイルと同一名のものがない事を確認する手段と、前記ファイルシステムが管理するファイル管理データの情報の中から、記録するのに必要なディスク容量を確保する手段を備え、また、前記ファイルシステムがファイル管理データを参照して記録するのに必要なディスク容量が確保できなかった場合、前記ファイルシステムは、前記データネットワークを介して接続されている、他のサーバ装置内部にあるファイルシステムに対してコンテンツファイルの情報の管理を要求する手段と、前記ディスク装置に対して書き込む際に、前記コンテンツファイルの情報の管理を要求した他のサーバ装置よりデータの書き込みに必要な情報を受け取る手段を備え、前記ディスク装置に対して書き込みを行うので、１つのサーバ装置に障害が発生した場合でも、該サーバ装置が管理するディスクエリアへのコンテンツファイルの記録処理は停止するものの、他のサーバ装置が管理するディスクエリアへの記録処理に関しては全く支障無く動作することが可能となる、という効果を得られる。

【００３６】また、請求項５および請求項６に係るディスク共有型分散サーバシステムによれば、請求項３または請求項４記載のディスク共有型分散サーバシステムにおいて、各サーバ装置内部のファイルシステムは、共有している前記ディスク装置に記録されている全コンテンツファイル名および前記コンテンツファイルの情報を管理しているサーバ装置を対応付けさせた管理テーブルを持ち、また、コンテンツファイルの追加または削除を行ったサーバ装置は、前記データネットワークを介して接続されている他のサーバ装置に対して、前記管理テーブルの更新を指示する手段を備えたので、コンテンツファイルの整合性を保ち、コンテンツファイル情報の検索を、接続されているサーバ装置全部に対して問い合わせすることなく、効率良くデータを検索することができる、という効果を得られる。

【００３７】また、請求項７および請求項８に係るディスク共有型分散サーバシステムによれば、請求項５または請求項６記載のディスク共有型分散サーバシステムにおいて、各サーバ装置内部のファイルシステムは、前記分割した記録領域における空き情報を空き領域テーブルとして持ち、また、前記空き領域情報の更新を行ったサーバ装置は、前記データネットワークを介して接続されている他のサーバ装置に対して、前記空き領域テーブルの更新を指示する手段を備えたので、ファイルシステム間における空き領域の整合性を保ち、接続されているサーバ装置全部に対して問い合わせをすることなく、効率良く空き領域の情報を検索することができる、という効果を得られる。



【００３８】また、請求項９に係るディスク共有型分散サーバシステムによれば、請求項３または請求項４記載のディスク共有型分散サーバシステムにおいて、システム初期化時に、前記ディスク装置における記録領域を論理的に分割する手段と、前記分割した記録領域に内在するコンテンツファイルの情報を取得する手段を備えたので、障害復旧時やサーバ装置およびディスク装置を増やすなどの場合に、各サーバ装置内部においてのファイルシステム情報の更新を容易にする、という効果を得られる。

【図面の簡単な説明】

【図１】本発明の実施の形態１による、ディスク共有型分散サーバシステムを示す構成図。

【図２】本発明の実施の形態１による、図１に示すディスク共有型分散サーバシステムに備えられるサーバ装置内部のファイルシステムが扱うファイル管理データの概要図。

【図３】本発明の実施の形態１による、ディスク共有型分散サーバシステムのデータ配信動作一例を示した図。

【図４】本発明の実施の形態１による、ディスク共有型分散サーバシステムのデータ配信動作一例を示した図。

【図５】本発明の実施の形態１による、図１に示すディスク共有型分散サーバシステムに備えられるサーバ装置内部のファイルシステムが扱うコンテンツテーブルの概要図。

【図６】本発明の実施の形態１による、ディスク共有型分散サーバシステムのデータ配信動作の一例を示した図。

【図７】本発明の実施の形態１による、ディスク共有型分散サーバシステムのデータ記録動作の一例を示した図。

【図８】本発明の実施の形態１による、ディスク共有型分散サーバシステムのデータ記録動作の一例を示した図。

【図９】本発明の実施の形態１による、ディスク共有型分散サーバシステムのデータ記録動作の一例を示した図。

【図１０】本発明の実施の形態２による、ディスク共有型分散サーバシステムを示す構成図。

【符号の説明】

１１ａ，１１ｂ，１１ｃ…サーバ装置

１２ａ，１２ｂ，１２ｃ…ファイル管理データ

１３…共有ディスク装置

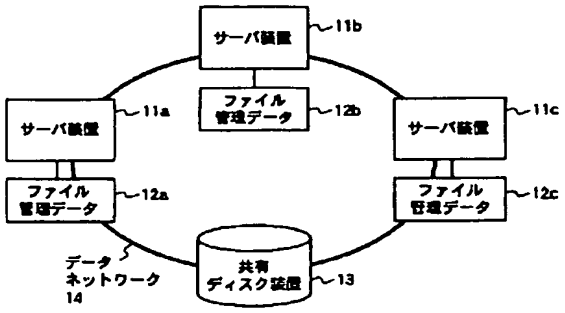
１４，１０２…データネットワーク

２１ａ，２１ｂ，２１ｃ…ディスク領域情報
２２ａ，２２ｂ，２２ｃ…コンテンツファイル情報
３１，３２，４１，４２，４３，４４，４５，６１，６２，６３，６４，７１，７２，７４，８１，８２，８３，８４，８５，８７，９１，９２，９３，９４，９６…制御信号

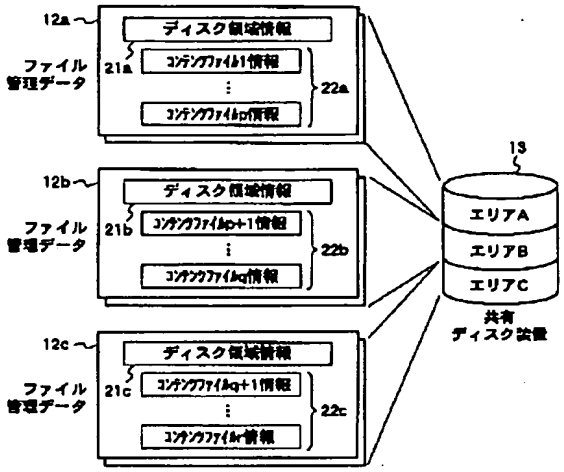


３３，３４，４６，４７，６５，６６，７３，７５，８６，８８，９５，９７…データの流れ
６０ａ，６０ｂ，６０ｃ…コンテンツテーブル
９０ａ，９０ｂ，９０ｃ…領域テーブル
１０１－１～１０１－ｎ…ディスクー体型サーバ装置

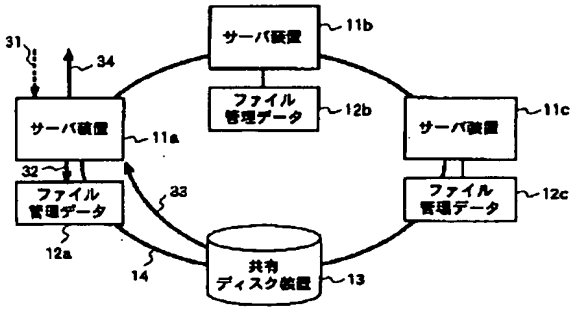
【図１】



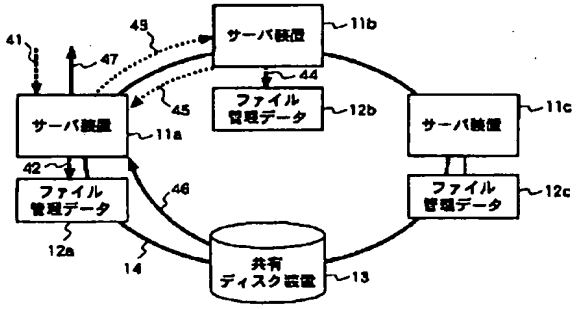

【図２】




【図３】



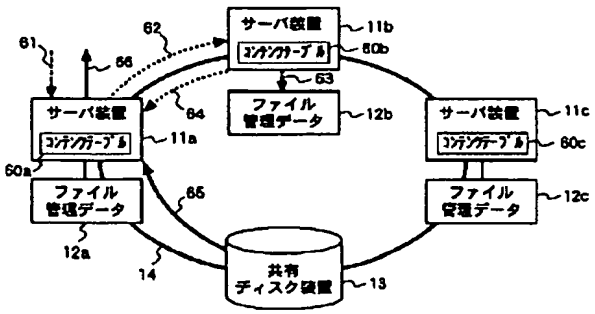

【図４】



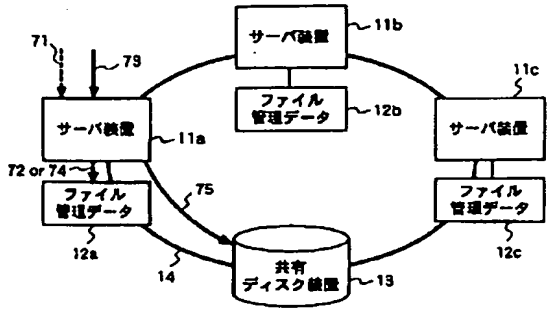

【図５】

| コンテンツファイル名 | 状態 | 保存場所 |
|---|---|---|
| コンテンツファイル1 | 未使用 | エリアA |
| ⋮ | | |
| コンテンツファイルp | 記録中 | |
| コンテンツファイルp+1 | | エリアB |
| ⋮ | | |
| コンテンツファイルq | | |
| コンテンツファイルq+1 | | エリアC |
| ⋮ | | |
| コンテンツファイルr | | |

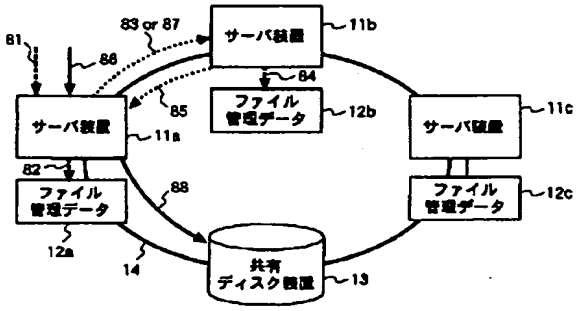
【図６】



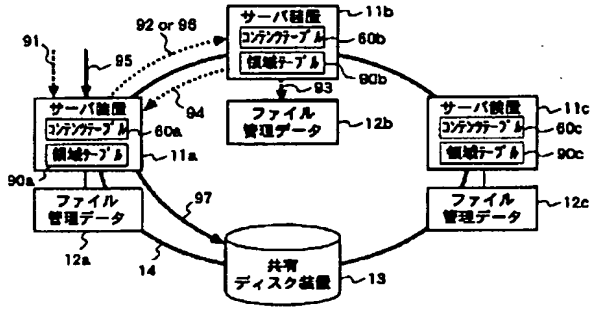
【図７】



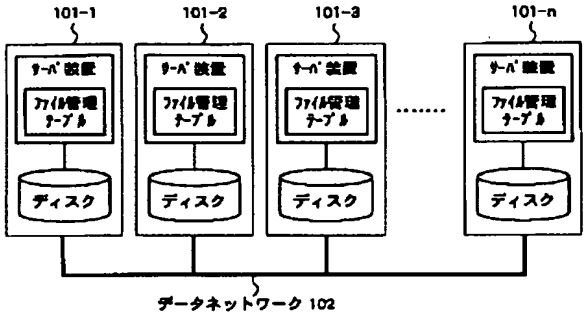

【図８】




【図９】




【図１０】




フロントページの続き

| (51) Int.Cl.[7] | 識別記号 | ＦＩ | ﾃｰﾏｺｰﾄﾞ(参考) |
|---|---|---|---|
| Ｇ０６Ｆ 15/167 | | Ｇ０６Ｆ 15/167 | Ｂ |